



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **Landesweite Normierung des Hessischen Kindersprachscreenings**

Tomasik, Martin J ; Lindner, Nina ; Weiland, Marina ; Schmidt, Melanie ; Moser, Urs ; Kieslich, Matthias

**Abstract:** Das Hessische Kindersprachscreening stellt ein Verfahren zur Feststellung des Sprachstands bei Kindern im Alter von 4;0 bis 4;6 Jahren dar. In der hier vorgestellten Überarbeitung wurden zunächst eine neue kriteriumsorientierte Skalierung und eine neue Normierung an einer flächendeckenden Stichprobe von insgesamt  $N = 43\,556$  Kindern aus knapp 900 Kindertageseinrichtungen in Hessen durchgeführt. Berichtet werden neben deskriptiven Item- und Skalenkennwerten auch die Befunde zur Objektivität, zur Reliabilität sowie zur Konstrukt- und faktoriellen Validität, die allesamt sehr zufriedenstellend ausfallen. Anschließend werden die diagnostischen Schwellenwerte (Cut-off-Werte) für die zwei Screeningergebnisse „sprachpädagogischer Förderbedarf“ und „medizinischer Abklärungsbedarf“ bestimmt und die Sensitivität und Spezifität dieser Schwellenwerte betrachtet. Es wird argumentiert, dass mit dem überarbeiteten Kindersprachscreening ein Verfahren vorliegt, dass sich durch insgesamt sehr zufriedenstellende psychometrische Eigenschaften auszeichnet und damit für die Zielsetzung geeignet ist, eine möglichst frühe und zielgerichtete Diagnose sprachlicher Defizite bei Vorschulkindern zu ermöglichen.

DOI: <https://doi.org/10.1026/0012-1924/a000246>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-197175>

Journal Article

Accepted Version

Originally published at:

Tomasik, Martin J; Lindner, Nina; Weiland, Marina; Schmidt, Melanie; Moser, Urs; Kieslich, Matthias (2020). Landesweite Normierung des Hessischen Kindersprachscreenings. *Diagnostica*, 66(3):178-189.

DOI: <https://doi.org/10.1026/0012-1924/a000246>

Sprache als wesentlicher Bestandteil des gesellschaftlichen Lebens und Instrument der Wissensvermittlung stellt einen wichtigen Baustein in der Regulation sozialer Beziehungen und der Repräsentation von Identität und Persönlichkeit dar. Die Erfassung und Förderung der Sprachkompetenz im vorschulischen Bereich gilt als elementare Grundlage der kindlichen Entwicklung und stellt die Schlüsselqualifikation des schulischen Erfolgs dar. „Eine bedarfsgerechte Sprachförderung muss möglichst früh ansetzen, um allen Kindern optimale Startchancen in ihre Schullaufbahn zu ermöglichen“ (Neugebauer & Becker-Mrotzek, 2013, S. 4). Als zentrales Themenfeld wurde bereits 2002 im Beschluss der 299. Kultusministerkonferenz festgesetzt, „die Sprachförderung – insbesondere die Förderung der sprachlichen Entwicklung von Kindern mit einer anderen Familiensprache als Deutsch – soll stärker als bisher in der pädagogischen Arbeit der Kindertageseinrichtungen und Vorschulklassen verankert und die Eltern dabei stärker einbezogen werden“ (Kultusministerkonferenz, 2002, S. 8).

Sowohl auf Bundes- als auch auf Länderebene wurden verschiedene Maßnahmen zur Etablierung von Verfahren zur Sprachstandsfeststellung sowie der Sprachförderung umgesetzt. Erklärtes gemeinsames Ziel war dabei, Sprachdefizite spätestens bis zur Einschulung auszugleichen. In den verschiedenen Bundesländern zeigt sich dabei aber eine große Heterogenität aufgrund unterschiedlicher länderspezifischer Traditionen, Richtlinien und Gesetzgebungen (Lisker, 2010; Schnieders & Komor, 2007). Eine durch das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache (Neugebauer & Becker-Mrotzek, 2013) bundesweite Analyse und Bewertung von Verfahren zur Sprachstandsfeststellung (Beobachtungsbögen, Einschätzverfahren, Testverfahren, Screenings) ergab, dass die Entwicklung von Verfahren zur Sprachstandsfeststellung auf einem guten Weg ist, dennoch großer Handlungsbedarf besteht. Anhand 32 definierter Qualitätsmerkmale zeigte sich, dass die Hälfte aller Verfahren 16 oder weniger Merkmale ausreichend erfüllte. An dieser Situation hat sich in den letzten Jahren wenig geändert. Die in einzelnen Bundesländern verwendeten Sprachscreenings variieren stark. Eine bundesweite Vereinheitlichung auf ein einendes Verfahren wäre diesbezüglich wünschenswert und bedürfte einer entsprechenden Koordinierungsstelle.

### **Hessisches Kindersprachscreening**

Verfahren zur vorschulischen Sprachstandsfeststellung und vorschulische Sprachförderkonzepte unterliegen einer hohen Veränderungsdynamik und unterscheiden sich

nicht nur bezüglich der inhaltlichen Umsetzung, sondern auch hinsichtlich der gesetzlichen Verpflichtung der Teilnahme (Lisker, 2010). Für das Land Hessen wurde der *Bildungs- und Erziehungsplan* als alters-, institutionen- und bildungsortübergreifendes bildungstheoretisches Konzept erarbeitet und als Maßnahmenkatalog umgesetzt, worin sprachliche Bildung als zentrale Aufgabe in der Elementar- und Primarpädagogik verortet ist (Fthenakis, Berwanger, & Reichert-Garschhammer, 2007). Hierfür wurde 2008 durch das Hessische Ministerium für *Soziales und Integration mit dem Kindersprachscreening (KiSS)* ein Instrument zur Sprachstandsfeststellung initiiert. Das KiSS wurde für den flächendeckenden Einsatz im Bundesland Hessen für mono- und multilinguale Kinder im Elementarbereich der Altersgruppe von 4;0 bis 4;6 Jahren entwickelt. Die Festlegung der Altersgruppe gewährleistet ein ausreichendes Zeitfenster, um im Bedarfsfall rechtzeitig vor Schuleintritt weitere diagnostische, therapeutische sowie fördernde Maßnahmen einleiten zu können. Das übergeordnete Ziel besteht darin, die Bildungschancen aller Kinder möglichst gleichberechtigt herzustellen.

**Beschreibung.** Das KiSS basiert auf Grundlage einer Revision des *Marburger Sprachscreenings* von Holler-Zittlau, Dux und Berger (2009; siehe auch Euler, Holler-Zittlau, Minnen, Sick, Dux, Zaretsky & Neumann, 2010). Die Struktur des Kinderbogens in der Version des KiSS 2.0 (Mertens, Dankwardt & Kislich, 2013) besteht aus insgesamt 46 Items, die auf acht Subtests verteilt sind. Die Items werden mit einem bis drei Punkten gewichtet. Die acht Subtests umfassen die Bereiche *Spontansprache*, *Sprachverständnis*, *Wortschatz* (Nomen, Adjektive), *Sprachproduktion*, *Aussprache*, *Grammatik* (mit den Aspekten Pluralbildung, Partizipbildung, Subjekt-Verb-Kongruenz, 2. Person Singular, Nebensatzbildung mit Konjunktion, Präposition im Akkusativkontext sowie Präposition im Dativkontext), *Nachsprechen von Wörtern* (wobei Kunstwörter zur Anwendung kommen) sowie *Nachsprechen von Sätzen*. Zu den Anforderungen der Untertests zählen beispielsweise im Subtest *Grammatik* die Nebensatzbildung mit Konjunktion, bei der das Kind Erklärungen oder Zweckbestimmungen für eine bestimmte Situation geben und diese sprachlich durch ein Satzgefüge mit der Finalkonjunktion „damit“ (Zweckbestimmung) konstruieren und einen Nebensatz mit Verb-Endstellung bilden soll. Im Bereich *Wortschatz* wird der aktive Wortschatz erfasst und im Bereich *Sprachverständnis* auch der passive, wobei der letztere Untertest so kurz ist, dass spezielle grammatische Strukturen hieraus nicht ableitbar sind. Im Bereich *Aussprache* wird der Wortschatz mit Absicht nicht geprüft, so dass das Wort vorgesprochen werden darf, wenn das Kind es nicht nennen kann. Die Bereiche

*Spontansprache* und *Sprachproduktion* sind unspezifisch und geben Hinweis auf die pragmatischen Fähigkeiten des Kindes.

**Durchführung.** Die Durchführung des Screenings erfolgt in der Kindertagesstätte durch die pädagogische Fachkraft, die zuvor durch einen KiSS-Sprachexperten geschult und über das Hessische Kindervorsorgezentrum zertifiziert wurde. Diese entscheidet auch nach eigenem Ermessen, ob die Deutschkenntnisse des Kindes für eine Durchführung des Screenings ausreichend sind oder nicht. Für den Fall, dass die Deutschkenntnisse nicht ausreichen sollten, erfolgt ein Screening ein halbes Jahr später, wobei in der Zwischenzeit eine informelle sprachpädagogische Förderung in der Kindertagesstätte stattfinden soll.

Zur Unterstützung der Durchführung dient eine Bildvorlage, auf die sich viele der Items beziehen und auf der die Kinder beispielsweise abgebildete Objekte benennen sollen. KiSS-Sprachexperten gehören den Berufsgruppen Logopädie, Sprachtherapie oder Sprachheilpädagogik an und durchlaufen eine zertifizierte Schulung im Hessischen Kindervorsorgezentrum. Als Multiplikatoren schulen KiSS-Sprachexperten die pädagogischen Fachkräfte vor Ort, stehen diesen beratend zur Verfügung, werten die Screeningbefunde aus und leiten diese an das zuständige Gesundheitsamt weiter.

Erfasst werden die individuellen kindlichen Sprachkompetenzen auf Grundlage von drei Elementen: dem Kitabogen (Angaben der pädagogischen Fachkräfte bezüglich der kindlichen Sprachentwicklung und Einflussfaktoren des Sprachumfeldes im Vorfeld des eigentlichen Sprachscreenings), dem Elternbogen (Angaben der Eltern bezüglich der kindlichen Sprachentwicklung und Einflussfaktoren des Sprachumfeldes im Vorfeld des eigentlichen Sprachscreenings) sowie dem Screeningbogen (Dokumentation des eigentlichen Sprachscreenings). Die Angaben der drei Elemente (Kitabogen, Elternbogen und Screeningbogen) werden anschließend im Kinderbogen zusammengefasst und bilden die sprachlichen Fähigkeiten und das Kommunikationsverhalten des Kindes ab.

**Auswertung.** Die erreichten Punktwerte werden zu einer Gesamtpunktzahl addiert. In der bisherigen Version des Kindersprachscreenings existiert ein definierter Cut-off-Wert für sprachpädagogischen Förderbedarf auf der Grundlage der erreichten Gesamtpunktzahl. Die maximal mögliche Gesamtpunktzahl des Kinderbogens betrug 105 Punkte als Summe aus den oben genannten acht Subtests. Der definierte diagnostische Schwellenwert für einen sprachpädagogischen Förderbedarf lag in der bisherigen Version des KiSS für mono- und multilinguale Kinder bei einer Gesamtpunktzahl von unter 50. Liegt die Gesamtpunktzahl bei 50 oder höher bei gleichzeitigem Vorliegen der Mindestanzahl in allen acht Subtests, resultiert

das Ergebnis *sprachlich unauffällig*, es sei denn, Kriterien für einen medizinischen Abklärungsbedarf (also Redeflussstörung, Stimmstörung, auditive Wahrnehmungsstörung, Hörstörung, k/g-Fehlbildung, r-Fehlbildung oder weitere Hinweise in den Fragebögen) werden erfüllt. Die inhaltliche Trennung zwischen sprachpädagogischem Förderbedarf und medizinischem Abklärungsbedarf besteht darin, dass Kinder mit sprachpädagogischem Förderbedarf ein Defizit aufweisen, das in der Form einer alltagsintegrierenden Sprachförderung in der Kindertagesstätte voraussichtlich behoben werden kann. In Abgrenzung dazu liefert das Ergebnis eines medizinischen Abklärungsbedarfs auf eine Störung (Sprachentwicklungs- oder Aussprachestörung, Redefluss- oder Stimmstörung, Hörstörung oder auditive Wahrnehmungsstörung), die sich voraussichtlich nicht in der Kindertagesstätte behoben werden kann. Diese Kinder sollten einem Kinderarzt vorgestellt und dann ggf. an eine logopädische Therapie überwiesen werden, weil davon auszugehen ist, dass die Störung nicht einfach durch mehr sprachlichen Input im Alltag behoben werden kann.

Die Berechnung der einzelnen Subtests erfolgte nach einem festgelegten Algorithmus, wonach Auffälligkeiten in einem oder mehreren Subtests zu einem sprachpädagogisch förderbedürftigen bzw. medizinisch abklärungsbedürftigen Ergebnis führen. Für multilinguale Kinder galten in den Subtests gesonderte Normen in Bezug auf den medizinischen Abklärungsbedarf. Sollten sich aus den Angaben des Kita-, Eltern- und Screeningbogens Hinweise auf eine Stimm-, Redefluss-, Hörstörung, auditive Wahrnehmungsstörung und/oder k/g- bzw. r-Fehlbildung ergeben haben, resultierte das Ergebnis eines medizinischen Abklärungsbedarfs, was eine Überprüfung durch den KiSS-Sprachexperten zur Folge hat. Die Auswahl der Störungen und Fehlbildungen erfolgte entweder aufgrund der Inzidenzraten oder aus theoretischen Überlegungen (Mertens et al., 2013). So gehören die k/g- bzw. r-Fehlbildungen zu den am häufigsten therapierten Lauten in der Aussprachetherapie. Hörstörungen und auditive Wahrnehmungsstörungen sind zentrale Risikofaktoren für die Sprachentwicklung und für erfolgreiches schulisches Lernen. Stottern und Stimmstörungen stellen weitere relevante logopädische Störungsbilder dar, die durch Sprachentwicklungsuntersuchungen in der Regel nicht direkt erfasst werden. Aus diesem Grund sollte diesen Hinweisen hier nachgegangen werden.

Weitere Punkte, die durch den Sprachexperten überprüft werden, sind eine unzureichende Aussprache, ein zusätzlicher Hinweis im Kita- oder Elternbogen sowie ein grenzwertiges Ergebnis. Neben dem errechneten Ergebnis wird zusätzlich immer eine Empfehlung durch den KiSS-Sprachexperten ausgesprochen, die vom Ergebnis abweichen

darf. Diese Empfehlung wird den Eltern im Elternbrief mitgeteilt. Des Weiteren besteht die Möglichkeit der Nachprüfung, beispielsweise bei zu geringen Deutschkenntnissen zum Zeitpunkt der Testung.

**Vorläufige Gütekriterien.** Mittels einer Stichprobe von  $N = 257$  Kindern im Alter von 4;0 bis 4;6 Jahren aus 43 Kindertagesstätten in Frankfurt am Main, Offenbach, Gießen, Korbach und dem Landkreis Fulda wurde die vorläufige Bestimmung der Gütekriterien vorgenommen (Euler et al., 2010). Die konkurrente Validierung erfolgte zunächst mit vier kombinierten Referenztests, nämlich mit dem *Reynell-Test (RDLS-III)*; Edwards, Fletcher, Garman, Hughes, Letts & Sinka, 1997) für den Bereich des rezeptiven Sprachverständnisses, mit der *Psycholinguistischen Analyse kindlicher Sprechstörungen (PLAKSS)*; Fox, 2005) zur Identifikation phonetischer und phonologischer Auffälligkeiten, mit fünf Untertests der *Patholinguistischen Diagnostik bei Sprachentwicklungsstörungen (PDSS)*; Kauschke & Siegmüller, 2002) zur Messung der syntaktisch-morphologischen Fähigkeiten sowie mit der revidierten Version des *Aktiven Wortschatztests für 3- bis 5-jährige Kinder (AWST-R)*; Kiese-Himmel, 2005) zur Bewertung der semantisch-lexikalischen Leistungen. Insgesamt war der konkurrente Validitätskoeffizient in Bezug auf diese vier Referenztests mit  $r = .85$  sehr hoch, wobei einschränkend beachtet werden muss, dass sich die Skalierung des Kindersprachscreenings gerade an diesen vier Referenztests orientiert hat, was zu einer Überschätzung des Validitätskoeffizienten geführt haben wird.

Die Punktwerte des Kindersprachscreenings wurden außerdem mit den zwei alternativen Screenings korreliert, nämlich mit dem *Heidelberger Auditivem Screening in der Einschulungsuntersuchung (HASE)*; Schöler & Brunner, 2003;  $r = .55$ ) sowie dem *Sprachscreening für das Vorschulalter (SSV)*; Grimm, 2003;  $r = .61$ ). Insgesamt ist die konkurrente Validität damit als sehr gut einzustufen.

Auch in der bereits oben erwähnten bundesweiten Analyse und Bewertung von Verfahren zur Sprachstandsfeststellung durch das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache erreichte das Kindersprachscreening eine gute und eine sehr gute Bewertung aufgrund der Berücksichtigung aller vier Basisqualifikationen, also der psychologischen, pädagogischen, sprachtherapeutischen und linguistischen Aspekte (Neugebauer & Becker-Mrotzek, 2013). Auch Kiese-Himmel und Rosenfeld (2012) sprechen dem Kindersprachscreening in einer vergleichenden Untersuchung von vier Untersuchungsverfahren eine positive Bewertung im Hinblick auf die psychometrischen Eigenschaften aus, bemängeln jedoch, dass das Verfahren nicht frei verfügbar sei.

**Probleme und gegenwärtige Weiterentwicklung.** Die bislang für die Bestimmung der psychometrischen Gütekriterien verwendete Stichprobe war sehr klein und selektiv, so dass auch keine sinnvolle Normierung vorgenommen werden konnte. Dieses Problem wurde nun durch die Verwendung einer großen und weniger selektiven Stichprobe angegangen.

Die bisherige Skalierung war außerdem nicht an einem Außenkriterium orientiert, sondern an den vier oben genannten Referenztests, was neben anderen Problemen vermutlich auch zu einer Überschätzung der Konstruktvalidität geführt hat. Neben einer Neuskalierung an dem Außenkriterium des sprachpädagogischen Förderbedarfs soll auch die Konstruktvalidität nun unter anderem an diesem Kriterium überprüft werden.

Auch konnte bisher die Sensitivität und Spezifität des Tests nicht direkt bestimmt werden, weil ein valides Außenkriterium gefehlt hat, so dass man sich auch hier mit einer indirekten Bestimmung dieser diagnostischen Kennzahlen an den Referenztests beholfen hat, was im Ergebnis aber nicht durchgängig zufriedenstellend war. Die Spezifität der im Screening als auffällig detektieren Kinder lag im Vergleich mit dem Expertenurteil bezüglich der vier Referenztests bei 78 %, die Sensitivität bei 88 %. Die Spezifität der Kinder mit dem Screeningergebnis des medizinischen Abklärungsbedarfs lag bei 93 %, die Sensitivität war aber unbefriedigend, da die Trennschärfe zwischen den Ergebnissen sprachpädagogisch förderbedürftig und medizinisch abklärungsbedürftig unzureichend war (Neumann, Holler-Zittlau, van Minnen, Sick, Zaretsky & Euler, 2011). Daraus abgeleitet wurde das Screeninginstrument um die zwei Subtests *Nachsprechen von Wörtern* zur Überprüfung des phonologischen Kurzzeitgedächtnisses sowie *Nachsprechen von Sätzen* zur Feststellung der zentralauditiven Verarbeitungsfähigkeit erweitert, deren psychometrische Gütekriterien hier zum ersten Mal überprüft werden. Wichtiger ist jedoch, dass nun die Bestimmung der Sensitivität und Spezifität an einem externen Expertenurteil erfolgen soll, was ein hochvalides und praktisch relevantes Außenkriterium darstellt.

**Zielsetzung.** Vor dem Hintergrund der bisherigen Entwicklung des Kindersprachscreenings und der vorläufigen Bestimmung der Gütekriterien sollen im Folgenden Daten aus einem flächenhaften Einsatz des Instruments in Hessen verwendet werden, um (a) eine neue Skalierung der Items aufgrund ihrer prognostischen Stärke vorzunehmen, (b) die sich so neu ergebende Verteilung der Punktwerte sowie die Prozentrangnormen in einer großen, landesweit repräsentativen Stichprobe von Kindern zu bestimmen, (c) weitere psychometrische Kennwerte im Hinblick auf die Reliabilität und die Validität der Skala zu ermitteln, (d) mögliche Effekte des Alters, des Geschlechts und des

sprachlichen Hintergrunds zu testen sowie (e) die im Hinblick auf Sensitivität und Spezifität optimalen diagnostischen Schwellenwerte für die beiden Screeningergebnisse des sprachpädagogischen Förderbedarfs sowie des medizinischen Abklärungsbedarfs zu definieren. Auf alle fünf Punkte wird im Folgenden, nach der Vorstellung der Stichprobe und des Untersuchungsdesigns, eingegangen.

## **Methode**

### **Stichprobe**

An der Untersuchung nahmen zwischen Januar 2010 und Mai 2017 insgesamt  $N = 43\,556$  Kinder aus knapp 900 Kindertageseinrichtungen in Hessen teil. Diese waren auf etwa ein Fünftel aller staatlich zugelassenen Einrichtungen in diesem Bundesland verteilt. Die Teilnahme einer Einrichtung an der Untersuchung und eines einzelnen Kinder in dieser Einrichtung an dem Screening wurde allen Einrichtungen empfohlen, erfolgte jedoch freiwillig. Von den teilnehmenden Kindern waren 51 % Jungen und 49 % Mädchen. Am Tag der Untersuchung waren sie zwischen 4;0 und 4;6 Jahre alt ( $M = 4.30$ ;  $SD = .15$ ). Das Alter nach ganzen Lebensmonaten war nicht gleichverteilt,  $\chi^2(6) = 3\,911.07$ ,  $p < .001$ , vielmehr waren vor allem die 4;4 sowie 4;5 Jahre alten Kinder überrepräsentiert und die 4;0 sowie 4;6 Jahre alten Kinder unterrepräsentiert. Bei 70 % der Kinder wurde kein Migrationshintergrund festgestellt. In 68 % der Fälle war die in der Familie gesprochene Sprache ausschließlich Deutsch.

Nicht für alle Kinder liegen alle personenbezogenen Angaben, Testwerte oder Empfehlungen der Sprachexperten vor, beispielsweise weil ein Test aus Gründen der fehlenden Mitarbeit oder wegen fehlender Deutschkenntnisse des Kindes nicht vollständig zu Ende geführt werden konnte. Somit stehen die Testwerte aufgrund der Einschätzung des Sprachexperten für  $N = 43\,486$  ( $> 99$  % der Gesamtstichprobe), Testwerte aufgrund der Einschätzung der Erzieherinnen für  $N = 43\,494$  ( $> 99$  %), Empfehlungen für einen sprachpädagogischen Förderbedarf für  $N = 40\,425$  (93 %) und Empfehlungen für medizinischen Abklärungsbedarf für  $N = 41\,417$  (95 %) Kinder zur Verfügung. Alle vier Angaben (Einschätzung des Sprachexperten, Einschätzung der Erzieherinnen und die beiden Empfehlungen) liegen für  $N = 40\,429$  (93 %) Kinder vor. Zwar haben Kinder, für die mindestens eine Empfehlung vorliegt, signifikant höhere Punktwerte als Kinder, für die die beiden Empfehlungen fehlen,  $F(2, 43,478) = 15.33$ ,  $p < .001$ , aber die Effektstärke ist hier mit  $\eta^2 = .001$  sehr gering. Insgesamt gesehen ist die Stichprobenselektivität also vernachlässigbar.



## **Prozedur**

Im Vorfeld des Sprachscreenings füllten die Eltern des Kindes den Elternbogen und die Bezugserzieherinnen den Kitabogen aus. Die Testdurchführung erfolgte durch vorher geschulte und zertifizierte Erzieherinnen in den Räumen der jeweiligen Kindertagesstätte und wurde auf dem Screeningbogen dokumentiert. Anschließend wurden die Angaben aus allen drei Bögen auf dem Kinderbogen zusammengefasst und an den Sprachexperten zur Kontrolle übergeben. Die Sprachexperten konnten die Angaben auf dem Kinderbogen gegebenenfalls korrigieren, werteten anschließend alle Informationen aus und kamen so zu einem algorithmisch festgelegten Ergebnis. Die Punktwerte, auf denen dieses Urteil beruht, werden im Folgenden für die Berechnung der psychometrischen und der diagnostischen Gütekriterien verwendet. Gleichzeitig haben die Sprachexperten, basierend auf dem Gesamteindruck, aber algorithmisch unabhängig von der Auswertung, eine eigene diagnostische Empfehlung in den Kategorien Kind unauffällig, sprachpädagogischer Förderbedarf, medizinischer Abklärungsbedarf sowie sprachpädagogischer Förderbedarf und medizinischer Abklärungsbedarf. Diese Empfehlung wird im Folgenden zur Validierung, zur Bestimmung der diagnostischen Schwellenwerte und zur Berechnung der Sensitivität und Spezifität des Tests verwendet.

## **Skalierung**

Von den 46 Items wurden bis auf Item 2 (wenn das Kind stottert) und Item 3 (wenn das Kind heiser ist) alle restlichen 44 für die Bestimmung der Gesamtpunktzahl verwendet. Den Items wurden unterschiedliche Punktwerte zwischen 1 und 3 Punkten (mit Ausnahme von 7 bzw. 8 Punkten bei den Items zum *Nachsprechen von Sätzen*) zugewiesen, womit sie bei der Bestimmung der Gesamtpunktzahl unterschiedlich gewichtet werden. Die Gewichtung erfolgte kriteriumsorientiert mit Blick auf das Screeningergebnis eines sprachpädagogischen Förderbedarfs. Die prognostisch stärksten bzw. trennschärfsten Items wurden mit 3 Punkten gewichtet und die weniger starken mit 2 bzw. 1 Punkt. Die maximal erreichbare Gesamtpunktzahl liegt damit bei 80 Punkten.

Die sechs Hinweise auf eine Stimm-, Redefluss-, Hörstörung, auditive Wahrnehmungsstörung und/oder k/g- bzw. r-Fehlbildung ergaben sich algorithmisch aus den Angaben in den drei Fragebögen. So wurde beispielsweise eine r-Fehlbildung automatisch diagnostiziert, wenn Item 23, Item 24 und Item 26 falsch gelöst wurden.

## **Demographische Angaben**

Das tagesgenaue Alter und das Geschlecht der Kinder wurden aus den Unterlagen der Kindertageseinrichtung bestimmt. Die zu Hause gesprochene Sprache zur Unterscheidung von monolingualen und multilingualen Kindern wurde aus dem Elternfragebogen übernommen.

## Ergebnisse

### Item- und Skalenkennwerte

In Tabelle 1 finden sich verschiedene Itemkennwerte für die einzelnen Items. Aus den Schwierigkeitsindizes  $P$  wird deutlich, dass die meisten Items sehr einfach zu lösen waren. Es sind aber auch 7 von 44 Items mit einem Schwierigkeitsindex von  $P < .50$  vertreten. Die Trennschärfen  $r_{it}$  für den Gesamttest und  $r_{is}$  für die einzelnen Subtests sind insgesamt als sehr zufriedenstellend zu bezeichnen. Jedes vierte Item weist eine Trennschärfe von  $r_{it} \geq .55$ , jedes zweite Item erreicht  $r_{it} \geq .43$  und drei von vier Items liegen bei  $r_{it} \geq .31$ . Die Trennschärfen  $r_{is}$  für die einzelnen Subtests liegen erwartungsgemäß noch ein wenig höher.

Der Mittelwert der Gesamtpunktzahl liegt bei  $M = 56.17$  ( $SD = 18.34$ ), wobei der Range der Skala voll ausgenutzt wird. Der Median liegt bei 62 Punkten, der Modalwert bei 71 Punkten. Die Dichteverteilung der erreichten Punktwerte ist in Abbildung 2 dargestellt. Diese ist deutlich linksschief ( $v = -.94$ ,  $SE_v = .01$ ) und eher steilgipflig ( $\gamma = .06$ ,  $SE_\gamma = .02$ ).

Die Verteilungseigenschaften der einzelnen Subtests finden sich in Tabelle 2. Wie beim Gesamttest sind auch bei den einzelnen Subtests die Punktwerte linksschief verteilt. Am deutlichsten zeigt sich das bei den Subtests *Spontansprache* und *Sprachproduktion*, am wenigsten ausgeprägt ist die Linksschiefe bei den Subtests *Nachsprechen von Wörtern* und *Grammatik*. Die Interkorrelationen der Subtests finden sich in Tabelle 3 oberhalb der Diagonalen. In absteigender Reihenfolge der Korrelationskoeffizienten korreliert der Gesamttest am stärksten mit *Grammatik* ( $r = .93$ ), dann mit *Wortschatz* ( $r = .87$ ), *Nachsprechen von Sätzen* ( $r = .85$ ), *Sprachverständnis* ( $r = .71$ ), *Nachsprechen von Wörtern* ( $r = .54$ ), *Aussprache* ( $r = .50$ ), *Sprachproduktion* ( $r = .32$ ) und schließlich mit *Spontansprache* ( $r = .22$ ).

### Normierung

In Tabelle 4 finden sich die Prozentrangnormen (in 5-Prozent-Schritten) für die Gesamtstichprobe. Hier wird noch einmal die Schiefe der Verteilung deutlich, die sich darin ausdrückt, dass die 25 % der Kinder mit den höchsten Punktwerten in einem Bereich von nur 10 Punkten streuen.

### Objektivität

Die Durchführbarkeit mit einheitlich geschulten zertifizierten pädagogischen Fachkräften in den Kindertagesstätten wurde bereits von Euler et al. (2010) ermittelt und bestätigt. Die Durchführungs- und Auswertungsobjektivität ist durch die regelmäßige Schulung und Zertifizierung der pädagogischen Fachkräfte und Sprachexperten sichergestellt. So erhalten sowohl die Sprachexperten als auch die Erzieherinnen zunächst eine etwa achtstündige Ersts Schulung mit einer theoretischen Einführung in das Kindersprachscreening und erproben die Durchführung und Auswertung an Fallbeispielen. In einer anschließenden praktischen Phase der Ersts Schulung führen die Erzieherinnen das Kindersprachscreening an mindestens einem Kind unter Supervision des Sprachexperten durch. Nach der Ersts Schulung finden einmal jährlich Nachschulungen statt, in denen Unsicherheiten, Neuerungen und häufige Fehlerquellen besprochen werden. Darüber hinaus werden alle Beteiligten angehalten, das Screening schriftlich zu dokumentieren, was zur Objektivität des gesamten Verfahrens beitragen soll.

### **Reliabilität**

**Interne Konsistenz.** Die interne Konsistenz des Gesamttests beläuft sich auf  $\alpha = .91$  nach Cronbach. Die Testhalbierungsreliabilität nach Guttman beträgt  $\rho = .95$ , wobei zur Testhalbierung Items mit gerader und ungerader laufender Nummer getrennt wurden. Beide Indikatoren für die Reliabilität sind nach Weise (1975) als hoch zu bezeichnen.

**Expertenübereinstimmung.** Die Übereinstimmung der Beurteilungen der Erzieherinnen mit denen der Experten findet sich, dargestellt als Kappakoeffizient ( $\kappa$ ) in der letzten Spalte von Tabelle 1. Die Sprachexperten haben die Angaben der Erzieherinnen übernommen und nur dann korrigiert, wenn offensichtliche Fehler vorgelegen haben. Durch dieses Vorgehen, das ja keine zwei unabhängigen Urteile miteinander vergleicht, ist die Übereinstimmung erwartungsgemäß sehr hoch bis nahezu perfekt. Lediglich die Items 35 ( $\kappa = .87$ ) und 36 ( $\kappa = .88$ ) weichen etwas nach unten ab, aber auch dort ist die Übereinstimmung trotzdem noch als hoch zu bezeichnen. Für den Gesamttest findet sich eine Übereinstimmung zwischen den Erzieherinnen- und Expertenbeurteilungen von  $\kappa = .70$  bzw.  $r = 1.00$ , was sehr gut ist. Für die einzelnen Subskalen zeigt sich ein ähnliches Bild. Selbstverständlich überschätzen diese Zahlen die tatsächliche Übereinstimmung und sind daher lediglich als Hinweis dafür zu verstehen, wie effektiv die Schulung den Erzieherinnen das Regelwerk des Kindersprachscreenings vermittelt hat.

### **Validität**

**Konstruktvalidität.** Ein konstruktvalides Screeningverfahren muss klar zwischen monolingual deutschsprachigen und multilingualen Kindern unterscheiden. Dieser Unterschied in Bezug auf den sprachlichen Hintergrund findet sich deutlich. Dabei wurden monolinguale Kinder, bei denen zu Hause ausschliesslich die Kindertagesstätte Deutsch gesprochen wurde, und multilinguale Kinder, bei denen zu Hause (mindestens) noch eine andere Sprache als die Kindertagesstätte Deutsch gesprochen wurde, unterschieden. Erwartungsgemäß lag der Mittelwert der multilingualen Kinder ( $M = 42.22$ ;  $SD = 19.71$ ) deutlich unter dem der monolingualen Kinder ( $M = 62.25$ ;  $SD = 13.85$ ), wobei die Effektstärke beträchtlich war,  $F(1, 43\ 361) = 14\ 623.68$ ,  $p < .001$ ,  $\eta^2 = .252$ .

Auch ist zu erwarten, dass ein konstruktvalides Verfahren sehr gut zwischen sprachlich unauffälligen Kindern und solchen, bei denen ein sprachpädagogischer Förderbedarf bzw. ein medizinischer Abklärungsbedarf besteht, unterscheidet. In Bezug auf den sprachpädagogischen Förderbedarf ist die Effektstärke sehr hoch,  $F(1, 40\ 389) = 34\ 614.16$ ,  $p < .001$ ,  $\eta^2 = .462$ . Als förderbedürftig diagnostizierte Kinder haben im Mittel deutlich niedrigere Werte auf dem Gesamttest ( $M = 34.23$ ;  $SD = 14.88$ ) als sprachlich unauffällige Kinder ( $M = 63.32$ ;  $SD = 12.94$ ). Die ROC-Analyse ergibt eine  $AUC = .926$ , was als sehr überzeugend bezeichnet werden kann. Für die Vorhersage des medizinischen Abklärungsbedarfs ist dagegen der Gesamttest alleine nur bedingt geeignet. Die Effektstärke (jeweils beurteilt nach Cohen, 1992) ist, auch nach Kontrolle des sprachlichen Hintergrunds, zwar signifikant, aber lediglich mittelhoch,  $F(1, 41\ 233) = 6\ 021.54$ ,  $p < .001$ ,  $\eta^2 = .127$ , und auch die Mittelwerte zwischen den medizinisch abklärungsbedürftigen ( $M = 42.11$ ;  $SD = 18.51$ ) und den unauffälligen Kindern ( $M = 58.76$ ;  $SD = 17.11$ ) unterscheiden sich nicht so stark wie zwischen den sprachpädagogisch förderbedürftigen und den diagnostisch unauffälligen Kindern. Die ROC-Analyse ergibt eine  $AUC = .757$ , was lediglich als ausreichend bezeichnet werden kann. Das ist ein Hinweis darauf, dass für eine Vorhersage des medizinischen Abklärungsbedarfs das Hinzuziehen weiterer Informationen aus den einzelnen Subtests notwendig ist. So ist beispielsweise allein der Zusammenhang zwischen dem medizinischen Abklärungsbedarf und dem Punktwert auf dem Subtest *Aussprache* mit  $F(1, 41\ 296) = 11\ 078.13$ ,  $p < .001$ ,  $\eta^2 = .212$ , deutlich höher.

**Faktorielle Validität.** Eine konfirmatorische Faktorenanalyse mit den sechs Subtests, welche durch mindestens zwei Items repräsentiert werden (*Sprachverständnis*, *Wortschatz*, *Aussprache*, *Grammatik*, *Nachsprechen von Wörtern* und *Nachsprechen von Sätzen*) sowie den zwei Einzelindikatoren, die als manifeste Variablen im Modell berücksichtigt wurden

(Spontansprache und Sprachproduktion) zeigt trotz der relativ hohen Zahl der manifesten Indikatoren eine zufriedenstellende Passung zu den Daten,  $\chi^2[876] = 77\,894.83$ ,  $p < .001$ ; RMSEA = .045; CFI = .935; TLI = .930, wobei wegen der dichotomen Indikatoren der WLSMV-Schätzer von Muthén (1993) verwendet wurde. Die standardisierten Ladungen liegen alle im Bereich  $.52 < \lambda < .93$  und alle Subtests weisen hochsignifikante Varianzen auf. Die latenten Interkorrelationen der Subtests finden sich in Tabelle 3, unterhalb der Diagonalen. Es finden sich dabei sehr hohe Korrelationen zwischen den Subtests *Sprachverständnis*, *Wortschatz* und *Grammatik*.

Eine genauere Untersuchung der Modifikationsindizes zeigte, dass sich die Passung des Modells erheblich verbessern ließe, wenn für einige Items Doppelladungen zugelassen würden. Dies gilt insbesondere für Item 19 und 21, welche auf allen anderen Subtests deutlich laden würden. Die zusätzlichen Ladungen würden dabei sowohl positiv als auch negativ ausfallen.

### **Effekte des Alters, des Geschlechts und des sprachlichen Hintergrunds**

Die Inspektion der Alters- und Geschlechtseffekte auf Itemebene (vgl. Tabelle 1) zeigt, dass es zwar signifikante Zusammenhänge gibt, die Effektstärken allerdings sehr klein sind. Hinzu kommt, dass lediglich die Zusammenhänge mit dem Geschlecht systematisch – meist zugunsten der Mädchen – sind, während die Alterseffekte in beide Richtungen zu gehen scheinen. Vor dem Hintergrund von Untersuchungen, die gezeigt haben, dass sich die sprachlichen Fähigkeiten von Jungen und Mädchen nur marginal zugunsten der Mädchen unterscheiden (z. B. Hyde & Linn, 1998; Leaper & Smith, 2004) und eingedenk dessen, dass der untersuchte Altersrange nicht sehr groß ist, sind diese Befunde wenig überraschend. Für die Gesamtskala zeigt sich ein ähnliches Bild. Während keine Alterseffekte nachweisbar sind ( $r = -.02$ ;  $p = .69$ ), liegt der Mittelwert der Mädchen ( $M = 57.38$ ;  $SD = 18.01$ ) leicht über dem der Jungen ( $M = 55.02$ ;  $SD = 18.57$ ), wobei die Effektstärke trotz statistischer Signifikanz minimal ist,  $F(1, 43\,481) = 180.45$ ,  $p < .001$ ,  $\eta^2 = .004$ . Aus diesem Grund wird auf eine alters- und geschlechtsspezifische Normierung verzichtet.

Auf den starken Effekt des sprachlichen Hintergrunds wurde bereits im Rahmen der Konstruktvalidität eingegangen. Trotz dieses starken Effekts wurde im Folgenden auf eine separate Normierung verzichtet, weil der geringere Sprachstand der multilingualen Kinder ja Teil des Phänomens ist, das man mit dem Screeningverfahren identifizieren möchte. Gleichwohl wurden, wie weiter unten deutlich wird, gruppenspezifische diagnostische Schwellenwerte für den medizinischen Abklärungsbedarf definiert, damit die multilingualen

Kinder nicht bloß aufgrund ihres bisher geringeren Kontakts mit der deutschen Sprache ein positives Screeningergebnis erhalten.

### **Bestimmung der diagnostischen Schwellenwerte**

Zur Bestimmung der im Hinblick auf Sensitivität und Spezifität optimalen Schwellenwerte braucht es ein Außenkriterium, was sozusagen den „wahren Zustand“ (*ground truth*) repräsentiert. In dieser Untersuchung bestand dieses Außenkriterium aus der Einschätzung des Sprachexperten, ob bei dem jeweiligen Kind sprachpädagogischer Förderbedarf oder sogar medizinischer Abklärungsbedarf vorlag. Zu dieser Einschätzung konnten die Testwerte des Kindersprachscreenings durch den Sprachexperten herangezogen werden, in der Entscheidung war dieser aber gänzlich frei und konnte immer ein eigenes diagnostisches Urteil abgeben. Nach diesem Urteil war bei 24 % der Kinder ein sprachpädagogischer Förderbedarf festzustellen, während ein medizinischer Abklärungsbedarf bei 14 % der monolingualen Kindern und 16 % der multilingualen Kindern vorlag. Vor diesem Hintergrund wurden für die beiden Diagnosekategorien die diagnostischen Schwellenwerte bestimmt.

**Schwellenwerte für den sprachpädagogischen Förderbedarf.** Zur Feststellung des sprachpädagogischen Förderbedarfs wurde zunächst der Gesamttestwert herangezogen, weil dieser mit einer  $AUC = .926$  prognostisch besonders stark war. Der Schwellenwert wurde hier bei 49.5 Punkten festgelegt, um sowohl die Sensitivität als auch die Spezifität des Screeningergebnisses zu maximieren. Daneben wurden die zwei Subtests mit der höchsten  $AUC$ , nämlich *Grammatik* und *Wortschatz*, herangezogen und zwar so, dass alleine durch das Unterschreiten auch nur eines der beiden Schwellenwerte ein positives Screeningergebnis ausgelöst wurde. Die beiden Schwellenwerte wurden so gewählt, dass die Spezifität auf Kosten ihrer Sensitivität maximiert wurde. Dieses Vorgehen ist deswegen zu rechtfertigen, weil davon ausgegangen werden kann, dass sehr viele auffällige Kinder bereits über den Gesamttest diagnostiziert werden. Alle Klassifikationskennzahlen finden sich in Tabelle 5. Gesamthaft wird eine Sensitivität und Spezifität von jeweils .87 erreicht, was für ein Screeningverfahren als sehr zufriedenstellend zu bezeichnen ist. Es werden 31 % aller Kinder als *sprachpädagogisch förderbedürftig* diagnostiziert, was 16 % der monolingualen und 61 % der multilingualen Kinder entspricht.

**Schwellenwerte für den medizinischen Abklärungsbedarf.** Für den medizinischen Abklärungsbedarf wurden die Schwellenwerte getrennt für die monolingualen und die multilingualen Kinder bestimmt, wobei hier wieder sowohl die beiden Subtests mit der

höchsten AUC als auch der Gesamttest zum Einsatz kamen. Anders als bei dem Ergebnis des sprachpädagogischen Förderbedarfs wurden jedoch die Schwellenwerte so bestimmt, dass sich die Spezifität auf den beiden Subtests und auf dem Gesamttest im Bereich von .95 bewegen sollte, um die Zahl der falsch-positiven Befunde zu minimieren. Die dadurch leidende Sensitivität wurde dadurch kompensiert, dass die Screeningergebnis des medizinischen Abklärungsbedarfs automatisch dann erfolgte, wenn sich bei dem Kind auch nur ein Hinweis auf mindestens eines von sechs spezifischen Störungsbildern (Redeflussstörung, Stimmstörung, auditive Wahrnehmungsstörung, Hörstörung, k/g-Fehlbildung oder r-Fehlbildung) zeigte. Diese Hinweise ergaben sich algorithmisch aus bestimmten Kombinationen von Merkmalen im Eltern-, Kita- und Screeningbogen und haben alle eine sehr hohe Sensitivität. In der Kombination der Verwendung von Hinweisen (mit hoher Sensitivität) und Testwerten (mit hoher Spezifität) konnte so die Klassifikationsgüte optimiert werden.

Für die monolingualen Kinder sind die Klassifikationskennzahlen in Tabelle 6 dargestellt. Der Schwellenwert auf dem Gesamttest wurde auf 34.5 Punkte festgelegt, bei der *Aussprache* waren es 6.5 Punkte und bei dem *Nachsprechen von Wörtern* 0.5 Punkte. Sowohl die Sensitivität als auch die Spezifität langen mit .87 in einem sehr zufriedenstellenden Bereich. Insgesamt erhielten 23 % der monolingualen Kinder das Ergebnis eines medizinischen Abklärungsbedarfs.

Bei den multilingualen Kindern wurde der Schwellenwert für den Gesamttest auf 14.5 Punkte, der Schwellenwert für die *Aussprache* auf 5.5 Punkte und der Schwellenwert für das *Nachsprechen von Wörtern* auf 0.5 Punkte festgelegt. In Kombination mit den Hinweisen ergab sich eine zufriedenstellende Sensitivität von .74 und eine sehr zufriedenstellende Spezifität von .85 bei einer Diagnoserate von 24 %. Alle Klassifikationskennzahlen für diese Untergruppe finden sich in Tabelle 7.

**Zusammenfassende Betrachtung.** Das Kindersprachscreening erlaubt auf der Grundlage der Testwerte und der Störungshinweise zwei diagnostische Entscheidungen, die durch die Wertung des Gesamttests natürlich nicht unabhängig voneinander sind. Während 17 % aller Kinder ausschließlich das Ergebnis eines sprachpädagogischen Förderbedarfs und 9 % der Kinder ausschließlich das Ergebnis eines medizinischen Abklärungsbedarfs erhalten, erhalten 14 % der Kinder beide Ergebnisse. Als völlig unauffällig werden demnach 60 % der Kinder eingestuft.

## Diskussion

Mit dem überarbeiteten und neu skalierten Kindersprachscreening liegt ein Verfahren zur Feststellung des Sprachstands im Vorschulbereich vor, das sich durch sehr gute psychometrische Gütekriterien, insbesondere im Hinblick auf das Ergebnis eines sprachpädagogischen Förderbedarfs, auszeichnet. Bereits die Vorgängerversion des Sprachscreenings hat in der Literatur positive Kritiken erhalten (Kiese-Himmel & Rosenfeld, 2012; Neugebauer & Becker-Mrotzek, 2013), auch wenn der diagnostische Wert, insbesondere im Hinblick auf das Ergebnis eines medizinischen Abklärungsbedarfs, nicht unhinterfragt blieb (Neumann et al., 2011). Die Überarbeitung umfasste insbesondere die Aufnahme von zwei Subtests (*Nachsprechen von Wörtern* sowie *Nachsprechen von Sätzen*), eine neue Skalierung (orientiert am Kriterium des sprachpädagogischen Förderbedarfs) sowie die Festlegung neuer diagnostischer Schwellenwerte. Die Aufnahme der zwei Subtests trug positiv zur Reliabilität und Validität des Verfahrens bei, wobei der Subtest *Nachsprechen von Wörtern* als diagnostisches Merkmal zur Vorhersage des *medizinischen Abklärungsbedarfs* dient. Die neue Skalierung verbessert insbesondere die diagnostische Güte des Verfahrens im Hinblick auf den sprachpädagogischen Förderbedarf. Mit der Festlegung der neuen diagnostischen Schwellenwerte ist es zudem gelungen, vor allem die als problematisch bekannte Sensitivität des Kindersprachscreenings (Neumann et al., 2011) zu erhöhen. Im Hinblick auf den sprachpädagogischen Förderbedarf und im Hinblick auf den medizinischen Abklärungsbedarf bei monolingualen Kindern ist die Sensitivität und die Spezifität (mit einem Wert von jeweils .87) als sehr zufriedenstellend zu bezeichnen. Auch im Hinblick auf den medizinischen Abklärungsbedarf bei multilingualen Kindern lässt sich die Spezifität (mit einem Wert von .85) als sehr zufriedenstellend bezeichnen. Einzig die Sensitivität des Verfahrens ist bei diesem Kriterium und dieser Untergruppe (mit einem Wert von .74) lediglich als zufriedenstellend zu bezeichnen. Die hier im Vergleich zur Spezifität relativ geringe Sensitivität wurde allerdings bewusst in Kauf genommen, um aus ethischen und pragmatischen Überlegungen den Anteil der falsch-positiven Ergebnisse in einem akzeptablen Rahmen zu halten.

Aus den so definierten Schwellenwerten ergeben sich Punktprävalenzen von 24 % für den sprachpädagogischen Förderbedarf und von 14 bzw. 16 % für den medizinischen Abklärungsbedarf. Diese Zahlen liegen im Rahmen dessen, was bisherige Studien zur Prävalenz von Sprachentwicklungsstörungen im Vorschulalter gezeigt haben, wobei die Raten in Abhängigkeit von der verwendeten Methode und der verwendeten Störungsdefinition zum



Teil erheblich schwanken. So schätzten Warnke und Niebergall (1993), dass etwa 10 % der Kindergartenkinder in ihrer sprachlichen Entwicklung verzögert sind. Mayr (1990) fand bei 15 % der 3- bis 6-Jährigen Sprachstörungen. Während die berichteten Prävalenzen bei Heinemann und Höpfner (1992), die im Rahmen einer Screeninguntersuchung bei 18 bis 34 % der Kinder Auffälligkeiten ermittelten, deutlich höher liegen, fanden Castell, Biener, Artner und Beck (1980) nur bei 8 % der Vorschulkinder einer behandlungsbedürftige Sprachstörung. Insgesamt spricht die Vergleichbarkeit der Prävalenzen für die Validität des vorliegenden Verfahrens

Auch die sonstigen psychometrischen Kennwerte und Merkmale sind insgesamt als sehr zufriedenstellend zu bezeichnen. Reliabilitäten über .90 erlauben eine präzise Feststellung des individuellen Sprachstands auch jenseits einer diagnostischen Anwendung. Die faktorielle Validität ist als zufriedenstellend zu bezeichnen und es zeigte sich, dass es gelungen ist, Items zu konstruieren, die den sehr hohen Anforderungen eines latenten Messmodells weitgehend zu genügen vermögen. Das Verfahren trennt außerdem sehr gut zwischen unauffälligen Kindern und solchen mit sprachpädagogischem Förderbedarf bzw. gut zwischen unauffälligen Kindern und Kindern mit medizinischem Abklärungsbedarf.

Eine Stärke dieser Untersuchung ist der Rückgriff auf eine große, flächendeckende Stichprobe, mit der sich die Verteilungen und Zusammenhänge auf der Populationsebene gut abbilden lassen. Auch das Vorhandensein eines validen Außenkriteriums für eine so große Anzahl von Untersuchungseinheiten und praktisch ohne Selektivitätseffekte ist positiv hervorzuheben. Einen kleinen methodologischen Nachteil stellt die Tatsache dar, dass Beurteilereffekte (beispielsweise im Sinne eine Milde-Streng-Dimension) nicht berücksichtigt werden konnten, was aber aufgrund der hohen Objektivität des Verfahrens nicht so sehr ins Gewicht fällt. Außerdem ist kritisch zu sehen, dass zwischen simultaner und sukzessiver Mehrsprachigkeit nicht unterschieden wurde, so dass hier keine differenzierten Aussagen möglich sind. Allerdings wurde die Kontaktdauer in den Fragebögen erfasst und stand so den Sprachexperten zur Einschätzung des Sprachstands vor dieser Hintergrundinformation zur Verfügung.

Insgesamt bietet das überarbeitete Kindersprachscreening den Praktikern in den Kindertageseinrichtungen ein psychometrisch solides Verfahren und stellt damit einen wichtigen Baustein zur Früherkennung von sprachlichen Defiziten bereits im Vorschulalter dar. Valide kann es aber nur sein, wenn aus der Diagnostik die richtigen Konsequenzen im Sinne einer sprachpädagogischen Förderung und/oder einer angemessenen medizinischen

Abklärung mit ggf. Behandlung oder Überweisung zur logopädischen Therapie gezogen werden (Kane, 2013). Wenn das geschieht, dann ist ein wichtiger Schritt zur Förderung der Chancengleichheit im staatlichen Bildungswesen gemacht.

### **Literaturverzeichnis**

- Castell, R., Biener, A., Artner, K & Beck, C. (1980). Artikulation und Sprachleistung bei drei- bis siebenjährigen Kindern: Ergebnisse der Untersuchung einer Zufallsstichprobe aus der Bevölkerung. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 29, 203–201.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Edwards, S., Fletcher, P., Garman, M., Hughes, A., Letts, C. & Sinka, I. (1997). *Reynell developmental language scales III: The University of Reading edition*. Windsor, Berkshire: NFER-Nelson.
- Euler, H. A., Holler-Zittlau, I., Minnen, S., Sick, U., Dux, W., Zaretsky, Y. & Neumann, K. (2010). Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder. *HNO*, 58, 1116–1123. doi:10.1007/s00106-010-2089-7
- Fox, A. (2005). *PLAKSS: Psycholinguistische Analyse kindlicher Sprechstörungen* (2. Aufl.). Frankfurt am Main: Harcourt Test Services.
- Fthenakis, W. E., Berwanger, D. & Reichert-Garschhammer, E. (2007). *Bildung von Anfang an: Bildungs- und Erziehungsplan für Kinder von 0 bis 10 Jahren in Hessen*. Wiesbaden: Hessisches Ministerium für Soziales und Integration und Hessisches Kultusministerium.
- Grimm, H. (2003). *SSV: Sprachscreening für das Vorschulalter: Kurzform des SETK 3-5*. Göttingen: Hogrefe.
- Heinemann, M. & Höpfner, C. (1992). Screening-Verfahren zur Erfassung von Sprachentwicklungsverzögerungen (SEV) im Alter von 3 ½ bis 4 Jahren. *Der Kinderarzt*, 23, 1635–1638.
- Holler-Zittlau, I., Dux, W. & Berger, R. (2009). *Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS): Ein Sprachprüfverfahren für Kindergarten und Schule* (5. Aufl.). Buxtehude: Persen Verlag.
- Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53–69. doi:10.1037/0033-2909.104.1.53

- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1–73. doi:10.1111/jedm.12000
- Kauschke, C. & Siegmüller, J. (2002). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen*. München: Urban & Fischer.
- Kiese-Himmel, C. (2005). *AWST-R: Aktiver Wortschatztest für 3- bis 5-jährige Kinder - Revision*. Göttingen: Hogrefe.
- Kiese-Himmel, C. & Rosenfeld, J. (2012). Analyse aktueller Untersuchungsinstrumente zur Früherkennung von Auffälligkeiten in Sprechen und Sprache in der pädiatrischen Vorsorgeuntersuchung U8. *Das Gesundheitswesen*, 74, 661–672. doi:10.1055/s-0032-1321724
- Kultusministerkonferenz (2002). *PISA 2000 – Zentrale Handlungsfelder: Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen*. Zugriff am 13.07.2018 unter [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2002/2002\\_10\\_07-Pisa-2000-Zentrale-Handlungsfelder.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2002/2002_10_07-Pisa-2000-Zentrale-Handlungsfelder.pdf)
- Leaper, C. & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, 40, 993–1027. doi:10.1037/0012-1649.40.6.993
- Lisker, A. (2010). *Sprachstandsfeststellung und Sprachförderung im Kindergarten und beim Übergang in die Schule: Expertise im Auftrag des Deutschen Jugendinstituts*. München: Deutsches Jugendinstitut. Zugriff am 29.11.2017 unter [https://www.dji.de/fileadmin/user\\_upload/bibs/expertise\\_sprachstandserhebung\\_lisker\\_2010.pdf](https://www.dji.de/fileadmin/user_upload/bibs/expertise_sprachstandserhebung_lisker_2010.pdf)
- Mayr, T. (1990). Zur Epidemiologie von Sprach-, Sprech- und Kommunikationsstörungen bei Vorschulkindern: Ergebnisse einer Screening-Untersuchung – Konsequenzen für die Gestaltung eines angemessenen Betreuungssystems. *Heilpädagogische Forschung*, 16, 14–20.
- Mertens, J., Dankwardt, S. & Kieslich, M. (2013). Das Hessische Kindersprachscreening (KiSS) – „medizinische Auffälligkeit“, Betrachtung von Einflussfaktoren. *Das Gesundheitswesen*, 75, P45.
- Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–243). Newbury Park, CA: Sage.

- Neugebauer, U. & Becker-Mrotzek, M. (2013). *Die Qualität von Sprachstandsverfahren im Elementarbereich: Eine Analyse und Bewertung*. Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Zugriff am 13.07.2018 unter [https://www.mercator-institut-sprachfoerderung.de/fileadmin/user\\_upload/Mercator-Institut\\_Qualitaet\\_Sprachstandsverfahren\\_Web\\_03.pdf](https://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Mercator-Institut_Qualitaet_Sprachstandsverfahren_Web_03.pdf)
- Neumann, K., Holler-Zittlau, I., van Minnen, S., Sick, U., Zaretsky, Y. & Euler, H. A. (2011). Katzungoldstandards in der Sprachstandserfassung: Sensitivität und Spezifität des Kindersprachscreenings (KiSS). *HNO*, 59, 97–109. doi: 10.1007/s00106-010-2231-6
- Schnieders, G. & Komor, A. (2007). Eine Synopse aktueller Verfahren der Sprachstandsfeststellung. In K. Ehlich (Hrsg.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund* (S. 261–342). Berlin: Bundesministerium für Bildung und Forschung. Zugriff am 30.11.2017. Verfügbar unter [https://www.bmbf.de/pub/bildungsforschung\\_band\\_11.pdf](https://www.bmbf.de/pub/bildungsforschung_band_11.pdf)
- Schöler, H. & Brunner, M. (2008). *Heidelberger Auditives Screening in der Einschulungsuntersuchung* (2. Auflage). Göttingen: Hogrefe.
- Warnke, A. & Niebergall, G. (1993). Entwicklungsdefizite: Klassifikation, Epidemiologie, Ätiologie und Prognose. In M. Döpfner & M. H. Schmidt (Hrsg.), *Kinderpsychiatrie* (S. 16–29). München: Quintessenz.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.